Statistical Predictions for Chain Ladder Data 14rit206

I. Mizera (University of Alberta, Edmonton),
M. Maciak (Czech University of Life Sciences, Prague),
M. Pešta (Charles University, Prague)

August 24, 2014 – August 31, 2014

The main purpose of our participation was to work out our idea of a new prediction proposal, which was experimentally showed to perform better than classical methods used in the actuarial science nowadays.

The objective of the BIRS "research in teams" proposal was to meet all three of us together to communicate recent developments in the area of chain ladder data prediction arising in the actuarial science and to make substantial progress towards finishing our common research in this topic, including a collaborative paper preparation.

We can undoubtedly say now that this objective was met. We found the BIRS environment very inspirational and beneficial for all of us: we were able to resolve most of our theoretical as well as practical issues involved, we set an ambitious course for our future collaboration and we did most of the work to finalize our paper and to submit it for publishing soon.

First of all, there was an acute demand to supply the new estimation proposal with some theoretical background that was still missing. There was also a need to verify the method via some simulation and real data results and we also wanted to provide a full scale comparison of all available methods. Moreover, we wanted to discuss and establish some common approaches for further developments as we can see a great potential for this type of statistical analysis in different areas, not just the actuarial science itself.

We believed our prediction proposal can perform significantly better than classical estimation procedures commonly used these days. We had some initial simulation results indicating such expectations however, no theoretical results were available mainly due to some highly technical issues involved in proofs. Although, we did not finish all the theoretical work we intended, it turned out our stay in Banff was very inspirational indeed - not only we showed our proposal can perform much better when using traditional goodness-of-fit measures but we also came up with a different understanding of the model background and we tried a new algorithm concept which even outperforms our own proposal while it requires much simpler and more straightforward theoretical calculations in general.

We also had many useful and bearing discussions together on possible extensions and improvements taking an advantage of some popular topics in statistics, like atomic pursuit methods, convex optimization or empirical processes, all of which can be further pursued and investigated in future.

1 Overview of the Field

The chain ladder data are common in the actuarial science: data comes in a triangular shape which arises in predicting reserves that insurance companies are required to keep to comply with regulatory statutes. There are several methods focusing on chain ladder data predictions proposed and even implemented in the statistical software environment R.

Development factors models are generally based on weighted linear regression models, e.g., *chain ladder* method [6]. These types of methods are very *sensitive* to the most recent claim amounts [9, Chapter 4], i.e., small changes in the last diagonal of the observed triangle produce large disturbances in the predictions. Moreover, these models are not robust against outliers [9, Chapter 2].

Bayesian methods in claims reserving combine expert knowledge or existing prior information with observations, e.g., *Bornhuetter-Ferguson* method [2], *Cape-Cod* method [3]. All of these classes of models require very disputable so-called expert judgement, which can be very influential and questionable as well in order what kind of expert knowledge should be take into account. *Generalized linear models* (GLM) in non-life insurance [4] suffer from the fact that the input data are in a triangular fashion, not a complete rectangular table. In many GLM reserving methods, there are too many parameters comparing to the number of observations and the number of parameters also depends on the number of observations. This fact yields very volatile parameter estimates together with not very robust prediction.

Several authors have already investigated *copula* modeling approach in chain ladder data, e.g., [8]. Generally, the *mean square error* (MSE) is widely used as one of the most important valuation criteria in the insurance business [5].

2 Scientific Progress Made in BIRS

Given the recent financial crisis and common corporate insurance policies the main emphasis in the chain ladder prediction is not only posed on overall precision in prediction but the final amount of reserves plays a key role as well and the lowest reserve possibly predicted is preferred.

While several methods, even implementations in the statistical software environment R, already exist, we were able to propose a new approach that outperforms all of them in both aspects above: better mean square error precision as well as smaller overall amount of reserves required. Our idea is motivated by the Delaigle and Hall (2013) paper but the theoretical background is however, different. Our prediction proposal is based on the ”nearest” foregoing period (two various proposals at the end depending on two different interpretation of the word ”nearest”) with a complete information rather than using a classical mean based approach, which seems to be quite unreliable and very non-robust mainly due to a very low amount of data (usually as low as 1 up to 10) used for prediction.

During our collaboration in BIRS, we were able to resolve some important theoretical issues we were dealing with as we set down the main principles for the methodological background of our proposal however, some technical proofs are quite challenging and we were not able to finish all necessary work in just one week.

On the other hand, while trying to solve some technical problems of the original proposal we pointed at another possible understanding of the principle behind and it turned out we could even outperform our original proposal while, in addition, having a great advantage of much simpler theoretical background and more straightforward technical proofs in hand with the alternative concept.

Thus, in BIRS we mutually compared both our proposals with standard chain ladder prediction approaches using over 300 real data scenarios - complete chain ladder data provided by different insurance companies where the upper triangle part of data is used for the prediction itself and the lower data triangle part was used to validate the prediction at the end.

Last but not least, we investigated the potential of the proposed methodology in some other areas of statistics and real life situations: specifically, we were provided with online newspaper data recording a post (article) specific amount of readers arising in time where the main interest was to predict further development in order to apply more targeted commercials for example. Another area with a similar data arising principle is the biometric theory of the so-called growth curves in Biology. Even though, there are some obvious dissimilarities and different theoretical background assumptions one has to account for, there is a great potential for the chain ladder prediction methods in statistics and there is no need one should be closely focused on actuarial science and actuarial data only.

3 Outcome of the Meeting

During the week at BIRS we had a great opportunity to meet together and to investigate and brainstorm all topics of our common interest, which we greatly appreciate. We proposed two variants for the chain ladder prediction method where both of them outperform classical approaches used nowadays, when evaluating the precision by the mean squared error (the only standard goodness-of-fit measure—sometimes even required by law—to be used by insurance companies).

While at BIRS, we were able to canvass and clear up most of major problems related to theoretical or practical issues and using a few hundreds of real data scenarios we were able to perform quite complex and overall comparison of classical methods with both our proposals, where both of them outperformed classical approaches significantly. We are now about to finish a paper on this topic concluding all important facts we learned together at BIRS. The paper will be submitted for publication in relevant scientific journal.

Additionally, we had a great chance to strengthen our collaboration as we can see far more potential for these methods in statistics and real life situations now. There are of course whole new problems and challenges arising but we believe we can bridge them over and exploit all hidden potential of our proposals and all their advantages. We all intend to continue with our collaboration in future and not only in the area of chain ladder prediction. This gives us confidence in possibly many more papers we could work on together.

References

[1] A. Delaigle and P. Hall, Classification using censored functional data, *JASA*, **108**, 1269–1283, (2013).

[2] R. L. Bornhuetter and R. E. Ferguson, The actuary and IBNR, *Proceedings of the Casualty Actuarial Society*, **LIX**, 181–195, (1972).

[3] H. Bühlmann, Estimation of IBNR reserves by the methods chain ladder, Cape Cod and complementary loss ratio, International Summer School 1983 (unpublished), (1983).

[4] P. D. England and R. J. Verrall, Stochastic claims reserving in general insurance (with discussion), *British Actuarial Journal*, **8**(3), 443–518, (2002).

[5] Š. Hudecová and M. Pešta, Modeling dependencies in claims reserving with GEE, *Insurance: Mathematics and Economics*, **53**(1), 786–794, (2013).

[6] T. Mack, Distribution-free calculation of the standard error of chain ladder reserve estimates, *ASTIN Bulletin*, **23**(2), 213–225, (1993).

[7] M. Pešta and Š. Hudecová, Asymptotic consistency and inconsistency of the chain ladder, *Insurance: Mathematics and Economics*, **51**(2), 472–479, (2012).

[8] M. Pešta and O. Okhrin, Conditional least squares and copulae in claims reserving for a single line of business, *Insurance: Mathematics and Economics*, **56**, 28–37, (2014).

[9] M. V. Wüthrich and M. Merz, *Stochastic claims reserving methods in insurance*, Wiley, (2008).